

How Much Does It Cost to Be a Scientist?

Benjamin Balsmeier^{a, b} and Maikel Pellens^{c, a}

a) *KU Leuven, Dept. of Managerial Economics, Strategy and Innovation, Leuven, Belgium*

b) *University of Münster, Institute for Organisational Economics, Münster, Germany*

c) *Centre for European Economic Research (ZEW), Mannheim, Germany*

Forthcoming in *Journal of Technology Transfer*

The final publication is available at Springer via <http://dx.doi.org/10.1007/s10961-014-9388-1>

Abstract

We examine the academe-industry wage gap. Once self-selection and different personal characteristics of academic and industrial scientists have been taken into account the wage gap narrows from 28% to 13%. The counterfactual wage faced by an academic scientist increases with time spent on development and decreases with time spent on research. This finding challenges the idea of a solely negative relationship between science and wages. We further find that preferences for science augment the relationship between research orientation and wages. Overall, the results have implications for policy makers that aim to increase development oriented research activities at universities, individual scientists thinking about whether to pursue a career in industry or academe, and managers trying to hire academic scientists.

JEL-Classification: G24, O31, 038

Keywords: Academe-Industry Wage Gap, Economics of Science, Matching

Contact details:

Benjamin Balsmeier

KU Leuven, Department of Managerial Economics, Strategy and Innovation

Naamsestraat 69, 3000 Leuven, Belgium

phone: +32 16 326 682, e-mail: Benjamin.Balsmeier@kuleuven.be

Maikel Pellens

Centre for European Economic Research

L 7,1, 68161 Mannheim, Germany

phone: +49 621 123 51 89, e-mail: Pellens@zew.de

* Balsmeier gratefully acknowledges financial support from the Flemish Science Foundation. Pellens gratefully acknowledges financial support from the National Bank of Belgium. Both authors thank participants in the LEI & BRICK workshop on the 'Organization, Economics, and Policy of Scientific Research', the Technology Transfer Society Conference, the DRUID Society Conference, a lunch seminar at KU Leuven, as well as Dirk Czarnitzki, Lee Fleming, Christoph Grimpe, Reinhilde Veugelers, Stijn Kelchtermans, Henry Sauermann, Toby Stuart, and Scott Stern for their insightful comments.

1. Introduction

It is a stylized fact that scientists earn more in industry than in academe. The National Science Foundation (NSF) Science and Engineering indicators show that the median wage of a doctorate recipient up to five years after graduation is \$85,000 in industry, compared to \$65,000 in academe (National Science Board, 2012). This difference is one of the main reasons that lead academic scientists to seek a position in industry (Walker, Vignoles, and Collins, 2010). Although monetary rewards represent only one component of the scientific reward system, they are highly valued by individual researchers. In a recent survey of more than 5,000 life scientists and physical scientists 37 percent of the academic scientists and 47 percent of the industrial scientists rated salary as a “very important” job factor (see Sauermann and Stephan, 2013, p. 897).

Despite the fact that the academe-industry wage differential plays a major role in scientists’ career decisions and, consequently, the production of knowledge in an economy, we know relatively little about the processes underlying the wage gap, or how the gap differs across individual scientists. Up to now, the debate centers on earning profiles within each sector, and the scientific orientation of industrial research.¹

A sparse literature tries to estimate the academe-industry wage difference. Agarwal and Ohyama (2013), for instance, compare wages of basic and applied scientists in industry and academe. While basic scientists earn less in academe than applied scientists, higher

¹ See for instance Sauermann & Roach (2014) or Stern (2004) for industrial wage dynamics. Many studies consider academic wage dynamics, including factors such as gender (McNabb and Wass, 1997; Barbezat, 1987; Bayer and Astin, 1968), seniority (Moore, Newman, and Turnbull, 1998), and international wage differences (Altbach, Reisberg, Yudkevich, Androushcak, and Pacheco, 2012; Stevens, 2004; Ong and Mitchell, 2000). An earlier stream of research estimates academics’ earnings functions, finding that scientists’ earnings are concave, peaking late in their career (Stephan, 1996; Creedy, 1988; Diamond, 1986a; Laitner and Stafford, 1985; Lillard and Weiss, 1979; Weiss and Lillard, 1978).

complementarities between applied and basic science lead to smaller wage differences within industry. It remains an open question, though, how much of the academe-industry wage gap is driven by self-selection in the scientific labor market, and how much is driven by researchers' focus on basic versus applied research. Some other studies examine the wage gap by comparing wages of academics with those of other workers (Walker et al., 2010; Stevens, 2004). While these studies provide useful information, they are limited in the assessment of academe-industry wage gaps because they compare wages of academics with those of non-academics (or, as in Walker et al., 2010, a selected subsample thereof).

We estimate the academe-industry wage gap controlling for self-selection into industry or academe and for differences in research focus. A rich dataset of Belgian Ph.D. holders allows us to match academic and industrial scientists, such that we can compare scientists with similar preferences, research activities, and other personal characteristics. Since we observe how much time scientists spend on research versus development, we can account in a highly detailed manner for differences in research activities. Other studies (e.g. Agarwal and Ohyama, 2013) have instead relied on a categorical classification of basic and applied research.

Once we correct for self-selection and differences in research activities, the average wage gap drops from 28% to 13%. Further, the size of the wage gap is partly determined by research activities: an academic who spends much time on research faces a significantly lower wage gap, while the wage gap is much higher for academics who spend much time on development.

This analysis complements Stern's (2004) and Sauermann and Roach's (2014) examinations of the science-wage relationship by showing that development activities relate positively to the wage gap. We are able to show this through detailed measures of the time scientists spend on research and development, whereas Stern (2004) uses relatively broad

categories, i.e. the chance to continue previous research and the permission to publish, in order to quantify a job's scientific environment. Sauermann and Roach (2014) estimate scientist's reservation wages for jobs with the opportunity to publish. We finally show that preferences have a moderating role: scientists with a high taste for science trade off monetary rewards against research orientation at a higher level than those scientists with a rather low taste for science.

These findings have important implications for scientists deciding to seek employment in industry or academe, managers trying to hire academic scientists, and policy makers aiming to increase applied research and commercialization at universities. A typical academic scientist, with a strong taste for science and a research-focused agenda, will earn significantly less in industry than a naïve average wage comparison would suggest. Moreover, the wage gap is by far the largest for academic researchers working on product developments. As academics are more and more urged to engage in development orientated research projects, careers in industry become more attractive, which in turn increases the risk of an academic brain drain. Scientists with a high taste for science, however, might still be better off in academe. Lastly, the significant remaining wage gap after matching indicates that factors other than preferences and research activities, e.g. differences in research valorization, are responsible for a large part of the academe-industry wage differential.

2. Conceptual framework

Conceptually, the wages of scientists are determined by their willingness to trade money for research-oriented jobs (preference effects), and whether there is sharing of the returns to new inventions between employers and scientists (productivity effects). The industry-academe wage gap is then the result of institutional differences in these factors between industry and academe.

To frame this discussion, we start with a brief description of differences in the scientific reward systems in academe and industry.

2.1. The scientific reward system in academe and industry

As mentioned above, wage differences between academe and industry are arguably driven by differences in the organization of research in both sectors (e.g. Sauermann and Stephan, 2013; Aghion, Dewatripont, and Stein, 2008). Academics tend to face flatter wage profiles than scientists in industry. This is especially the case in European countries where academics are hired as civil servants. Wages in academe are further largely independent of scientific productivity, because academic research is difficult to monitor, and because performance contingent pay risks encouraging scientists to focus on safe projects with guaranteed results (Konrad and Pfeffer, 1990).² Instead of money, the academic reward system depends in large part on nonmonetary rewards to incentivize their researchers to engage in the most fruitful research projects. Nonmonetary rewards of science include the intellectual challenge involved with new projects, pleasure from “puzzle solving”, and peer recognition by other scholars (Aghion, Dewatripont, and Stein, 2008; Stephan, 1996; Dasgupta and David, 1994; Merton, 1973).

Firms, in contrast, are supposed to engage in science to maximize profits, focusing on applied research topics with the highest expected returns for new product developments (Arora and Gambardella, 1994; Rosenberg, 1990; Cohen and Levinthal, 1990, 1989). Industrial researchers thus tend to work on more narrowly defined research projects with emphasis on the commercial potential of their outputs.

² Some academics do manage to increase their income with consulting, speaking fees, or prizes.

Stern (2004) shows that these differences in the scientific reward systems have important implications for the relationship between science and wages in industry. On the one hand, firms could pay out a compensating differential, because researchers in industry have less freedom in research than they would have in academe (which Stern calls the preference hypothesis). This leads to a negative relation between research orientation and wages in industry (similar results have been found by Sauermann and Roach, 2014; Agarwal and Ohyama, 2013; Sauermann and Stephan, 2013; Fini and Lacetera, 2010; Lacetera, 2009; and Aghion, Dewatripont, and Stein, 2008).

On the other hand, firms might participate in science to increase their absorptive capacity, capture spillovers and raise their R&D productivity. This might lead to a positive relation between research and wages if firms share at least a part of those productivity gains with their scientists. Stern labels this the productivity hypothesis (see also Cohen and Levinthal, 1990; Rosenberg, 1990). The academe-industry wage gap varies depending on which effect dominates equilibrium wages of scientists in industry.

2.2. Scientists' wages

Our conceptualization and argumentation regarding the relationships between wages, scientific orientation, and tastes of researchers follows the logic of Stern's (2004) economic model of science and wages.³ We briefly illustrate our thoughts against this background and incorporate them in an extended model below. The original model starts with defining the utility function of an individual scientist as the combination of his wage and the pleasure derived from working at a science-orientated workplace:

³ For brevity, we only show those parts of Stern's model that help to clarify our thoughts and argumentation. For details and all the implicit assumptions made to keep the model simple we refer to Stern (2004) and the cited literature there.

$$U_i = \lambda_0 + \alpha_s \gamma_i SCl_j + w_i,$$

where α_s represents the scientist's preference for science, γ_i is the scientist's ability, SCI indicates whether a job has a scientific orientation or not, and w_i is the scientist's wage.

Stern then defines the profit generated by scientist i and earned by firm j as:

$$\pi_{i,j} = \gamma_i(\beta_0 + \beta_s SCl_j) - w_{i,j} - \delta SCl_j.$$

Firms receive returns β_s if they adopt a scientific orientation, but they have to pay a fee, δ , for e.g. specialized equipment and materials needed to carry out research projects. They also have to pay the wage of the scientist, regardless of their engagement in science. As scientists are assumed to receive a share $\phi \in (0,1)$ of the returns to their research, the equilibrium wage industrial scientists earn is then derived as:

$$\omega_{i,j}^* = \gamma_i \beta_0 + \gamma_i(\phi \beta_s - \alpha_s) SCl_j.$$

The wage equation describes the relationship between the productivity of research, the preference for science and the scientific orientation of the workplace. The preference effect dominates the productivity effect as long as α_s is larger than the quasi-rent stemming from research that a scientist can extract ($\phi \beta_s$).

Without changing the basic setup of the model, we assume that a job can have a certain degree of scientific orientation, which is determined by the time a scientist can spend on research as well as on development. We assume that the time spent on research serves a scientist's preference for science relatively more than the time spent on product development. Incorporating this notion in the utility function of the scientist gives us:

$$U_i = \lambda_0 + \alpha_s \gamma_i R_j + \frac{\alpha_s}{\theta} \gamma_i D_j + w_i,$$

where R_j represents the time a scientist spends on research, and D_j the time he spends on development. The parameter $\theta > 1$ embodies the idea that a scientist's preference for research is less served by focusing on product or process developments as compared to general research.

The profit function of a firm is correspondingly determined by the returns to research and development activities. We argue that the returns to research as well as development are increasing with decreasing marginal returns, but development activities are expected to yield generally higher returns than research activities. One reason might be the higher risks to achieve commercially viable outputs by engaging in basic research as opposed to development oriented research. This gives us the following profit equation:

$$\pi_{i,j} = \gamma_i(\beta_0 + \beta_r^z R_j + \beta_d^y D_j) - w_{i,j} - \delta(R_j + D_j),$$

where β_r^z represents the returns to research activities and β_d^y represents the returns to development activities, with $z \in (0,1)$, $y \in (0,1)$ and $z < y$.⁴ The equilibrium wage is then given by:

$$\omega_{i,j}^* = \gamma_i \beta_0 + \gamma_i [(\phi \beta_r^z - \alpha_s) R_j + (\phi \beta_d^y - \frac{\alpha_s}{\theta}) D_j].$$

This wage equation describes the relationship between the time scientists spend on research and development, their preference for science and the productivity of their activities. The wages of scientists decrease in the time spent on research as long as the preference for science, α_s , is larger than the quasi-rent from research activities a scientist can extract ($\phi \beta_r^z$).

Regarding the time spent on development, the compensating differential, $\frac{\alpha_s}{\theta}$, is relatively low, while the productivity effect, $\phi \beta_d^y$, is relatively large. Hence, it is more likely that the

⁴ As such, the profit function cannot explain why some firms choose to engage only in research. For simplicity reasons we refrained from incorporating this in the model, but it could, for instance, be realized by accounting for different functional forms of the costs of research and development activities. As noted above, the assumption of higher productivity of development is not essential for the model, but does help us to derive implications. In practice both activities are complementary and face diminishing returns, and which activity has the highest marginal productivity depends on the relative levels of investment in both.

productivity effect dominates the preference effect regarding development oriented research. In this case, scientists' wages are expected to increase with the time spent on development.

We can also see that a high taste for science (large α_s) will increase the negative wage effect of the time spent on research as long as the returns to research are rather low. In contrast, as long as the productivity effect of development oriented research activities dominates the preference effect, taste for science will, if at all, only slightly negatively moderate the positive productivity effect of development oriented research (cf. section 2.4 below).

2.3. The wage gap

Drawing on our extended model of Stern (2004), we can derive implications for the academe-industry wage gap. We build on the basic assumption that in academe spending more time on research does not directly affect scientists' wages,⁵ while in industry our conceptualization and the extant evidence indicates a negative relationship. This leads us to formulate:

Hypothesis 1: The wage gap faced by an academic scientist decreases with the proportion of time spent on research.

It is not straightforward to derive a similar hypothesis regarding how academic scientists would profit from conducting development oriented research relative to their counterparts in industry. On the one hand, academic scientists might profit more, since they can typically appropriate a part of the returns to their inventions. On the other hand, they might profit less, since universities have less assets, knowledge, and incentives to successfully commercialize innovations. This leads us to formulate the following two hypotheses regarding the effect of development oriented research activities on the wage gap:

⁵ However, academic scientists who successfully focus on research could have higher chances for promotion, and hence increase earnings in the long run.

Hypothesis 2a: The wage gap faced by an academic scientist increases with the proportion of time spent on development.

Hypothesis 2b: The wage gap faced by an academic scientist decreases with the proportion of time spent on development.

While these hypotheses have direct implications for the average academe-industry wage differential, the actual size of the gap depends on the marginal impacts of spending time on research or development, and the difference between the average time academic and industrial scientists actually spend on research or development. Therefore, we refrain from formulating hypotheses about whether the wage gap will become wider or narrower once these differences of academic and industrial scientists have been accounted for. A detailed evaluation follows in section 5.

Two extant empirical investigations are closely related to the present paper's examination. Stern's (2004) empirical analysis of young biologists' job offers revealed a negative relation between the scientific orientation of a particular job and the offered wage. He concluded that the preference effect dominates potential productivity effects in industry. In line with this finding, Sauermann and Roach (2014) have recently shown that scientists with a high taste for science demand higher salaries if they have to give up the right to publish research findings. Both results back hypothesis 1.⁶ We are able to go beyond Stern (2004) and Sauermann and Roach (2014) by observing the time scientists actually spend on 'research' and 'development' in their jobs. This allows us to overcome the focus on specific aspects of 'science' formerly employed in the literature. Specifically, we will distinguish between the proportion of time a scientist spends on

⁶ It is worthwhile to mention that these findings do not indicate that there are no productivity effects at all, but rather that the preference effect on average dominates potential productivity effects with regard to basic research.

research activities, which are not directly related to product or process development, and the proportion of time a scientist spends on the development of new products and processes.

2.4. Scientists' preferences

The academe-industry wage differential can also be explained by the preferences of scientists. Specifically, how much compensation an industrial scientist demands for working in a more or less scientifically orientated job depends critically on his 'taste for science' (Sauermann and Roach, 2014). Researchers who have a high taste for science are considered to have a high intrinsic motivation to perform research, a desire for intellectual challenge, and strong preferences for classic aspects of the academic system, e.g. freedom in choosing research topics, focus on basic research questions, and rewards through peer recognition (Sauermann and Roach, 2014; Roach and Sauermann, 2010; Stern, 2004; Stephan and Levin, 1992; Merton, 1973).⁷ As taste for science rises, scientists derive higher utility from doing research, leading them to accept lower compensation for a stronger research orientation of their jobs. We formulate:

Hypothesis 3: Provided that the wage gap an academic scientist faces becomes smaller with more time spent on research (Hypothesis 1), this effect is larger for scientists with a high taste for science.

Moreover, a well-established literature has shown that scientists' choices for a career in academe or industry are affected by their taste for science (Sauermann and Roach, 2012; Roach and Sauermann, 2010; Stern 2004). This leads to strong self-selection in the scientific labor market which has to be taken into account in wage comparisons. Selection effects in the labor

⁷ Scientists' taste for business, on the other hand, represents motivation through extrinsic factors, such as wages, extralegal benefits, and job security. Noteworthy, both preferences are not mutually exclusive, meaning that scientists can have a high taste for science as well as a high taste for business (Sauermann & Roach, 2012; Agarwal & Ohyama, 2012; Roach & Sauermann, 2010). We include the researchers' taste for business into our empirical analysis to account for potential confounding effects.

market are comprehensively revealed in Roach and Sauermann (2010), who found that young Ph.D. students with a high taste for science are more likely to prefer a career in academe, while those with a greater concern for salary, access to resources, and applied research prefer a career in industry (see also Agarwal and Ohyama, 2013).

3. Methodological remarks

One of the fundamental problems in the empirical assessment of the academe-industry wage gap is thus self-selection in the scientific labor market, which may distort a consistent estimation of the factual wage differential and its determinants. Since scientists are not randomly assigned to industry or academe but rather choose to do so based on their personal characteristics and preference for science, standard OLS regressions are likely to yield inconsistent estimates.

As non-random assignment is a very common problem in empirical policy evaluations, a number of methods have been developed to address this issue (Heckman, 1976; see Blundell and Costa Dias (2009) for an overview of methods used in labor market policy evaluations, or Imbens and Wooldridge (2009)). In this application, we control for selection through a matching estimator.⁸ Matching techniques have been used by several labor economists interested in wage gap estimations and labor policy evaluations (see, among others, Nopo, 2008; Frolich, 2007; Gerfin and Lechner, 2002; Angrist, 1998; Card and Sullivan, 1988). Matching has several advantages over parametric estimators (cf. Imbens and Wooldridge, 2009). It aims to create the

⁸ Several other methods are available to deal with selection, including difference-in-difference designs, regression discontinuity designs, instrumental variable estimation, and control functions. The difference-in-difference method requires panel data on scientists that switch from academe to industry (or the other way around) due to an exogenous shock. As the data we employ is a cross-section and no exogenous shock is observed we cannot apply this estimator. A regression discontinuity design is also not applicable because there is no (arbitrary) rule that would determine the treatment (being an academic scientist). IV estimators as well as selection models rely on the idea to instrument selection into industry or academe through an instrumental variable, which affects the treatment but not the outcome, i.e. the wage gap. It is notoriously difficult to find a valid instrument in this particular case and we could not find a suitable instrument for this application either.

conditions of an experiment with random assignment by constructing a control sample with equal characteristics as the treatment sample. While it does not impose a functional form on the selection, it relies on the assumption that there are no unobserved variables that drive selection into academic or industrial positions. In order to minimize concerns in this regard we show below that our control sample consists only of industrial scientists that are indistinguishable from academic scientists along all known drivers of self-selection and several other personal characteristics.

The goal of our analysis is to calculate how much money an academic scientist would have earned in an industrial position where she spends the same proportion of her time on research and development oriented activities. Technically, we want to calculate the average treatment effect on the treated:

$$E(\alpha_{TT}) = E(Y_T|S = 1) - E(Y_C|S = 1),$$

where S denotes whether the individual has been treated (1: in academia), Y_T denotes the wage of the treated group (academic scientists), and Y_C represents the wages of the control group (industrial researchers). Since we cannot directly observe the wages of academic researchers had they chosen to work in industry, we have to estimate them. This is made possible by Rubin's Conditional Independence Assumption, which states that for a scientist with a set of exogenous characteristics X participation and potential outcomes are independent from one another (Rubin, 1977):

$$E(Y_C|S = 1, X) = E(Y_C|S = 0, X)$$

As long as we are able to find for all academic scientists a counterpart in industry ($0 < P(S = 1|X) < 1$), which is known as the common support restriction, we can use the wages of industrial scientists to estimate the wages of the academic scientists had they not been treated:

$$E(\alpha_{TT}) = E(Y_T|S = 1, X = x) - E(Y_C|S = 0, X = x).$$

In order to condition the samples on observable characteristics, we employ nearest neighbor matching (Heckman, Ichimura, and Todd, 1998) where we choose the most similar control observation for each treatment observation. This is however difficult to achieve when many dimensions at the same time are considered. It is therefore more convenient to use propensity score matching, where the matching occurs on a propensity score $P(X)$ estimated using a Probit model. This has no clear advantages or disadvantages over multidimensional matching (Heckman, Ichimura, and Todd, 1998; Rosenbaum and Rubin, 1983), but makes it much easier to find adequate controls. According to the classic literature on propensity score matching, the average treatment effect on the treated can be calculated as the mean difference of the matched samples:

$$\hat{\alpha}_{TT} = \frac{1}{n^T} (\sum_i Y_i^T - \sum_i \widehat{Y}_i^C),$$

with \widehat{Y}_i^C being the counterfactual wage for scientist i and n^T being the sample size of the academic scientists. Below, we will compare the descriptive academe-industry wage gap, calculated as the difference in means, with (1) standard OLS estimates of the wage gap, (2) the difference in means after balancing the samples of academic and industrial scientists using propensity score matching, and (3) OLS estimates of the wage gap based on the balanced sample of academic and industrial scientists. If selection effects in the scientific labor market are present, we would expect that the descriptive wage gap is the largest, while OLS estimates based on the balanced sample after matching show the lowest wage gap.

4. Data

We employ data from the Belgian edition of the Careers of Doctorate Holders (CDH) survey (Federaal Wetenschapsbeleid, 2006). The survey took place in 2006 (reference year: 2005), and was created by the OECD in cooperation with Eurostat and the UNESCO Institute for Statistics. The primary goal was to gain detailed knowledge about career paths of doctorate holders.⁹ The CDH survey contains detailed questions, including selected ones about the scientists' research activities, which allow for a much more accurate approximation of activities than other large scale data sources (e.g. the frequently used National Science Foundation's Survey of Doctorate Recipients). Combined with information on scientists' and researchers' earnings, it offers a unique chance to investigate the academe-industry wage differential and how the nature of research activities affects it.

The response rate of the survey is estimated at 17.7 percent (7,160 responses). Of these, we select natural scientists and engineers currently employed in industry or academe and being in the first 30 years of their careers. After removing observations with inconclusive answers (see appendix 1 for a detailed overview of the sample construction), and retaining only those who answered all relevant questions, we arrive at a final sample covering 486 academic scientists and 759 industrial scientists. Below, we describe all variables used in the analysis (see Table 1 for variable definitions). Table 3 lists corresponding summary statistics. Table A3 in the appendix presents pair-wise correlations.

⁹ Detailed information on the project can be found in Unesco (2012), Eurostat (2012), or OECD (2013). Summary statistics based on the full survey have been published in Auriol (2007, 2010). Statistics on the Belgian data collection can be found in Moortgat & Van Mellaert (2011). For the methodological background and core questionnaire see Auriol et al. (2012).

Table 1: Variable definitions

Variable	Description
<i>WAGE</i>	Reported earnings before taxes in 2005, taking all occupations into account.
Motives	
<i>TASTE_SCIENCE</i>	Respondents were asked to indicate why they chose a research career (7 possible reasons in binary form). We applied exploratory factor analysis (results reported in Appendix 2), extracting two factors.
<i>TASTE_BUSINESS</i>	<i>TASTE_SCIENCE</i> correlates positively with intellectual challenge, independence, contribution to society, and work circumstances, representing intrinsic motivation and higher preference for ‘classic’ science. <i>TASTE_BUSINESS</i> correlates positively with salary, extralegal benefits, career prospects, and job security, representing extrinsic motivation.
Nature of research activities	
<i>RES_SHARE</i>	Respondents were asked how much of their time they spend on average on research and development activities. <i>RES_SHARE</i> is the sum of the share of time assigned to “executing or guiding research” and “interpreting others’ research”. <i>DEV_SHARE</i> is the sum of the share of time assigned to “improving existing products or processes” and “developing new products or processes”. Both range from 0 to 100 percent of time.
<i>DEV_SHARE</i>	
General controls	
<i>CAREER_YEARS</i>	Time since graduation in years.
<i>AGE</i>	The age of the respondent in years.
<i>TIME_CURRENTJOB</i>	The time the respondent has been employed in the current job, in years.
<i>RELATIONSHIP</i>	1 if the respondent indicates to be cohabitating or married, 0 otherwise.
<i>CHILDREN</i>	1 if the respondent indicates to have children younger than 18, 0 otherwise.
<i>GENDER</i>	1 if the respondent is female, 0 otherwise.
<i>NATURAL SCIENCES</i>	1 if the respondent indicates to be active in the natural sciences (“mathematics, computer science, informatics”; physics; chemistry; geology; biology; other natural sciences), 0 if the respondents indicates to be active in engineering or technological sciences (civil, mechanical, chemical, electrical, electronic engineering; material sciences; medical engineering techniques; environmental science; biotechnology; nanotechnology; other engineering or technological disciplines).
<i>HOURS_WORKED</i>	Self-reported average number of hours worked per week.
Controls for ability	
<i>TIMETOPHD_PROP</i>	The time the respondent spent to complete his or her doctorate. Calculated as the deviation from the mean time to completion within detailed scientific field and ten-year graduation cohort. A score equal to one indicates that the doctorate was completed at the average time within field and cohort. A score lower than one indicates that the doctorate was completed more quickly.
<i>SCHOLARSHIP</i>	1 if the respondent received a government or private grant to fund his or her doctorate, 0 otherwise.

Dependent variable: wage

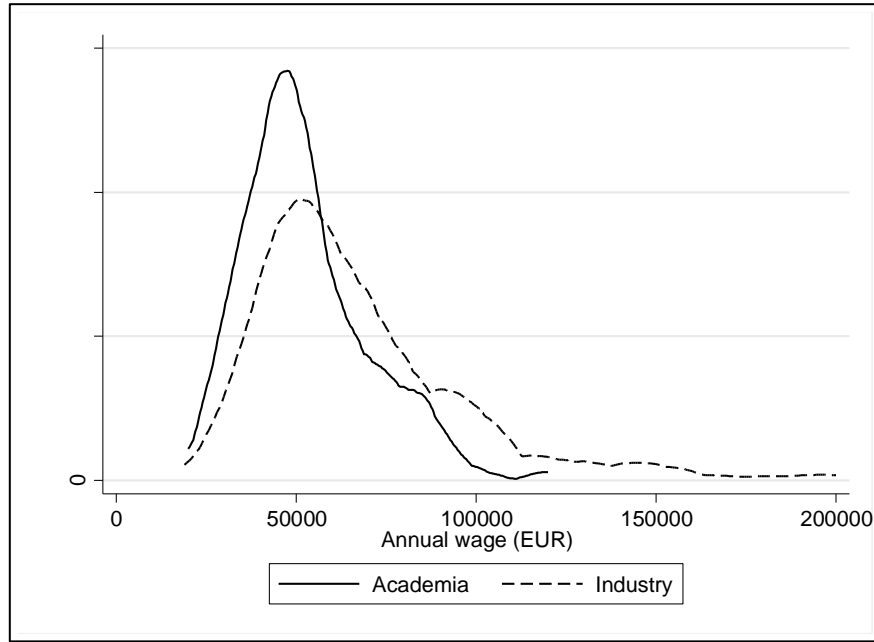
The outcome of interest is the sum of the scientist's or researcher's self-reported annual earnings before taxes.^{10,11} Academics reported earning 52,203 EUR on average, while industrial researchers reported 67,019 EUR. This amounts to a wage gap of 28.3 percent. Median wages are 49,100 EUR for academics versus 60,000 EUR for industrial researchers. These numbers are very similar to the median wages reported by the Survey of Doctorate Recipients in the US, which were 65,000 USD for academics versus 85,000 USD for industrial researchers, or a corresponding wage gap of 30.8% (National Science Board, 2012). Figure 2 plots the wage distributions within academe and industry. The earnings of industrial scientists are shifted somewhat to the right compared to those of academic scientists. Industrial wages also show a significantly longer right tail than academic wages, reflecting flatter earnings profiles in academia.

Academic wages in Belgium differ structurally from those in the United States: academics are usually hired as civil servants and the corresponding wage scales for public sector workers are set by a governmental decree in negotiation with strong labor unions. Wages in the Belgian academic sector are therefore only to a small degree influenced by competition in the scientific labor market. The main source of variance in academic wages stems from job tenure, position and side activities, e.g. speaking fees, which are included in the CDH's definition of the yearly wage.

¹⁰ The results change only slightly when we take hourly wage as the main variable of interest instead of the annual wage. The corresponding results are presented in Appendix 4.

¹¹ In order to ensure a clear differentiation between academic and industrial scientists and control for systematic differences caused through multiple employments, we discarded scientists and researchers with multiple jobs.

Figure 1: Wage distribution by sector



Explanatory variables

Activities: time spent on research and development

The CDH survey includes questions about how scientists spend their working time. These questions allow a more fine-grained measurement of scientists' activities compared to most statistics used in the extant literature, which typically indicate the main focus of a scientists' research activities, e.g. basic versus applied research, but do not provide a weighted measure of different activities (e.g. Agarwal and Ohyama, 2013; Sauermann and Stephan, 2013).

In order to approximate the proportion of time a scientist spends on research activities (*RES_SHARE*) we sum the share of time spent on “performing or guiding research” and “interpreting the research of others”. The time a researcher spends on development oriented research (*DEV_SHARE*) is proxied by the total share of time spent on “improving existing products

or processes” and “developing new products or processes”.¹² Descriptive statistics show that academic scientists spend on average 50 percent of their time on research activities, and 7 percent of their time on development activities, while these numbers correspond to respectively 29 percent and 40 percent for industrial researchers. The picture that these numbers draw is in line with the common notion that academic scientists tend to focus more on basic rather than applied research, while industrial scientists tend to spend much more time on product and process development. Hence, we expect the wage gap to narrow after balancing the data for these factors: industrial researchers who spend more time on research and less time on development than the average researcher should earn less in industry, as hypothesized in our conceptual framework.

A crucial advantage of our data is the multidimensional continuous measurement of engagement in research and development. Many scientists spend nontrivial shares ($\geq 1\%$) of their time on research as well as development: 30% of the academics, and 65% of researchers. Only 16% of academics indicate to be full-time researchers, and only 6% of industrial researchers are engaged full-time in development.

Motives

We proxy scientists’ preferences through their motives as captured through the survey question “*why did you choose a research career?*” The answers were recorded in binary form. Table 2 shows the differences between industry and academe, which are in line with previous research on motivation-driven sector choice based on US scientists’ data (Roach and Sauermann, 2010).

¹² Agarwal and Ohyama (2013) define scientists’ activities as primarily basic or primarily applied in nature, while Sauermann and Stephan (2013) make a distinction between scientists primarily engaged in basic research, those primarily engaged in applied research, and those primarily engaged in development. Our measure of research activities is most likely mainly driven by basic research activities but also picks up applied research components.

Table 2: Motives by sector

	Academia	Industry	<i>p</i> -value of two-sided <i>t</i> -test on mean differences
<i>INTELLECTUAL CHALLENGE</i>	0.84	0.62	$p < 0.0001$
<i>SALARY</i>	0.04	0.08	$p = 0.0106$
<i>EXTRALEGAL BENEFITS</i>	0.00	0.02	$p = 0.0035$
<i>CAREER</i>	0.16	0.18	$p = 0.3046$
<i>JOB SECURITY</i>	0.09	0.04	$p = 0.0039$
<i>WORK CIRCUMSTANCES</i>	0.29	0.18	$p < 0.0001$
<i>INDEPENDENCE</i>	0.65	0.33	$p < 0.0001$
<i>CONTRIBUTION TO SOCIETY</i>	0.17	0.06	$p < 0.0001$

Notes: Share of respondents in each sector who indicated being motivated by the respective factor.

The literature on scientists' motives finds that scientists have a certain taste for science as well as business (Roach and Sauermann, 2010; Sauermann and Roach, 2012; Agarwal and Ohyama, 2013). Taste for business represents motivation through extrinsic factors, such as wages, extralegal benefits, and job security (for a definition of taste for science see section 2 and the cited literature). In order to enhance comparability with this literature, we extracted tastes for science and business from the answers through exploratory factor analysis (see Appendix 2 for detailed results).¹³ In line with the extant literature, academic scientists score higher on taste for science than industrial scientists (0.43 versus -0.28, two-sided *t*-test on mean differences: $t(1243) = -13.07$, $p < 0.001$), and lower on taste for business (-0.12 versus 0.79, two-sided *t*-test on mean differences: $t(1243) = 3.52$, $p < 0.001$). Because of this, we expect the wage gap to narrow after balancing the sample through matching: scientists in industry who have a high taste for science and a low taste for business are more willing to trade off wage for publishing opportunities, and should hence earn less than average (Sauermann and Roach, 2014).

¹³ Our results are robust to alternatively taking the sums of *CHALLENGE*, *CONTRIBUTION*, and *INDEPENDENCE* to represent taste for science.

It should be noted that scientists and researchers can have a high taste for science and business at the same time (cf. Sauermann and Roach, 2012), and that tastes might change through time or socialization in the workplace (Sauermann & Roach, 2012), while we measure them at the time the survey was carried out. As such, our results concerning tastes should be interpreted with care.

As mentioned above, matching estimates need to account for all factors that affect selection into industry or academe. Therefore, we include as many factors as possible in the analysis. In addition to tastes and research activities, we include scientists' ability, effort, career experience and demographic characteristics as covariates. It might be argued that some of these factors do not directly cause selection, or are subject to reverse causality. For example, scientists potentially choose the amount of effort they put into their work as a function of wage prospects, and time spent on research and development are consequences of selection instead of causes. As such, the matching should not be viewed with a causal interpretation in mind. Nevertheless, the identification of the wage gap should only improve by conditioning on these factors, so we include all of them as covariates.

Ability

Ability influences our estimations in two ways. First, more able scientists are likely to be more productive, and thus earn more. Second, ability has been shown to cause selection into applied or basic research: in academe, more able scientists are more likely to prefer basic research topics than applied research, since academe tends to offer more resources to conduct basic research. Higher complementarities between basic and applied research in industry cause these differences in access to resources to be smaller, leading to no such selection in industry (Agarwal and Ohyama, 2013). As ability is usually unobservable to the econometrician it is

important to find suitable proxies that pick up as much as possible of the unobserved variance. Common ability proxies such as education are not applicable since all respondents attained a Ph.D. Indicators such as publication or citation counts are also biased in the context of this study: institutional differences between industry and academe lead to academics being more prone to publish. Reliance on publications would thus lead to an upward ability bias in academics' ability as compared to industrial researchers.

Nevertheless, the CDH survey allows us to create two proxies that are not affected by selection into industry or academe. The first one is based on the main source of funding of the respondents' Ph.D. studies. Government and private Ph.D. scholarships tend to select strongly on applicants' previous academic performance.¹⁴ At the same time, Ph.D. students are strongly encouraged to apply for these scholarships. Therefore, having a scholarship (*SCHOLARSHIP*) should separate a higher-skilled group from a lower-skilled group. Agarwal and Ohyama (2013) employ grant data in a similar fashion as one among two other controls for ability.

The time a Ph.D. holder spent to complete his dissertation serves as a second proxy for ability. In order to deal with heterogeneity across fields and time, we calculate time to completion as a proportional deviation from the mean time to completion within a field and a ten-year graduation cohort (*TIMETOPHD_PROP*).¹⁵ A value equal to one indicates that the scientist completed the Ph.D. in an average amount of time, while values less than one indicate faster and thus better performance. A similar measure is used by Agarwal and Ohyama (2013), where time to completion of Bachelor's degree is taken.

¹⁴ For instance, one of the main Flemish institutions distributing Ph.D. scholarships (FWO) lists "research ability and potential (including course results)" and "research skills and methodology" as the first two selection criteria for its Ph.D. fellowships (FWO, 2013).

¹⁵ We consider the researchers' detailed research domain (listed in table 1) for this calculation.

Both proxies have the advantage of being based on the respondent's Ph.D. education, and are thus independent of later career decisions. The main disadvantage is that there is some unobserved ability variance that we have to leave uncontrolled. We find some difference in ability proxies between the academic and industrial samples: industrial researchers do not on average complete their dissertation faster or slower than academic researchers, but are slightly more likely to obtain a Ph.D. scholarship (56% versus 50%). Hence, we expect that balancing on ability proxies will slightly narrow the wage gap.

Effort

Effort is an important determinant of wage. Therefore, we include the self-reported number of hours worked during an average week as a further matching variable (*HOURS_WORKED*).¹⁶ Academic scientists report working 50 hours per week on average, while industrial researchers report 48. Since more working hours positively correlate with pay, the wage gap should slightly lower when this difference is accounted for.

Career Experience

Experience is another well-known driver of scientists' wages: several studies indicate that academic scientists' earnings profiles are concave from below, and peak late in their careers. (Stephan, 1996; Creedy, 1988; Diamond, 1986; Laitner and Stafford, 1985; Lillard and Weiss, 1979; Weiss and Lillard, 1978). Age and cohort effects are also important for scientists' productivity (Stephan, 1996; Stephan and Levin, 1992), further affecting wages. In the matching, we control for three time-related factors. The first, *AGE*, is the natural age of the scientist. The second, *CAREER_YEARS*, is the time since the scientist obtained his Ph.D. The last time-related measure is the time spent in the current job (*TIME_CURRENTJOB*). As scientists spend more time in

¹⁶ Appendix 4 provides results based on the hourly wage instead of the annual wage as an alternative to control for differences in scientists' efforts.

a position, they accumulate job-specific skills which probably increase wage. In academe specifically, wages are known to rise with job tenure and academic rank, which both increase over time.

The academic and industrial samples are balanced in terms of career age (11.85 years in academe vs. 11.45 years in industry, $p=0.381$), but the scientists in the academic sample are slightly older than those in industry (41.70 years versus 40.88 years, $p=0.080$). They have also spent more time in their present position (8.84 years versus 7.41, $p<0.001$). Because industrial researchers are on average younger and have less experience, the wage gap should widen once the samples have been balanced in this dimension.

Demographics

Supplementary to the work related attributes, we take some demographic factors into account that may have an influence on wage setting as well as career decisions. These include gender (*GENDER*, takes value one if the scientist is a woman), relationship status (*RELATIONSHIP*, takes value one if the scientist is married or cohabiting), and whether the scientist has children (*CHILDREN*, takes value one if the scientist has children younger than 18). Scientists in academe are more likely to be female (22% versus 16%, $p=0.011$), are not more likely to be in a relationship (83% vs. 85%, $p=0.436$), but are less likely to have children (54% versus 64%, $p<0.001$). If the gender wage gap is larger in industry than in academe, we would expect the industry-academe wage gap to narrow once women are equally represented in the industrial sample. We refrain from formulation expectations regarding the effect of balancing on children, as it is not clear ex ante what the effects could be.

Field

Finally, we include a dummy variable that indicates whether a scientist is active in natural sciences. The baseline category represents engineering. Natural scientists are significantly more likely to work in academe than in industry (63% versus 51%, $p < 0.001$). Assuming that they work on more basic research projects compared to engineers, the wage gap should again narrow once we take this into account.

As most characteristics of scientists that are supposed to be positively (negatively) correlated with wages are overrepresented (underrepresented) in the sample of industrial researchers compared to their academic counterparts, we expect that the wage gap between academe and industry narrows significantly after balancing both samples.

Table 3: Summary statistics

Variable	Academia before matching (n=486)					Potential controls in industry (n=759)					<i>p</i> -value of two-sided <i>t</i> -test on mean differences
	Mean	Std. Dev.	Min	Median	Max	Mean	Std. Dev.	Min	Median	Max	
<i>TASTE_SCIENCE</i>	0.43	0.86	-1.63	1.02	1.27	-0.28	0.98	-1.94	-0.36	1.47	<i>p</i> <0.001
<i>RES_SHARE</i>	50.34	35.67	0.00	50.00	100.00	29.18	28.98	0.00	20.00	100.00	<i>p</i> <0.001
<i>DEV_SHARE</i>	7.43	15.63	0.00	0.00	100.00	40.16	32.33	0.00	40.00	100.00	<i>p</i> <0.001
<i>TASTE_BUSINESS</i>	-0.12	0.43	-0.55	-0.29	2.05	0.08	1.23	-0.55	-0.14	9.60	<i>p</i> <0.001
<i>CAREER_YEARS</i>	11.85	8.38	0.00	10.50	29.75	11.45	7.53	0.00	10.00	29.92	<i>p</i> =0.381
<i>AGE</i>	41.70	8.75	27.00	41.00	63.00	40.88	7.48	27.00	39.00	61.00	<i>p</i> =0.080
<i>TIME_CURRENTJOB</i>	8.84	9.23	0.00	5.00	37.00	7.41	6.77	0.00	5.00	30.00	<i>p</i> =0.002
<i>RELATIONSHIP</i>	0.83	0.38	0.00	1.00	1.00	0.85	0.36	0.00	1.00	1.00	<i>p</i> =0.436
<i>CHILDREN</i>	0.54	0.50	0.00	1.00	1.00	0.64	0.48	0.00	1.00	1.00	<i>p</i> <0.001
<i>GENDER</i>	0.22	0.41	0.00	0.00	1.00	0.16	0.38	0.00	0.00	1.00	<i>p</i> =0.011
<i>TIMETOPHD_PROP</i>	1.00	0.30	0.05	0.96	2.40	0.97	0.28	0.02	0.94	2.72	<i>p</i> =0.176
<i>SCHOLARSHIP</i>	0.50	0.50	0.00	0.00	1.00	0.56	0.50	0.00	1.00	1.00	<i>p</i> =0.023
<i>HOURS_WORKED</i>	50.10	8.38	38.00	50.00	72.00	48.11	7.03	38.00	48.00	72.00	<i>p</i> <0.001
<i>NATURAL SCIENCES</i>	0.63	0.48	0.00	1.00	1.00	0.51	0.50	0.00	1.00	1.00	<i>p</i> <0.001
<i>WAGE</i>	52203.12	17348.76	20000.00	49100.00	120000.00	67019.13	29556.26	19000.00	60000.00	200000.00	<i>p</i> <0.001

Notes: Definitions of the variables are provided in Table 1.

5. Results

5.1. The size of the wage gap

Before tackling potential selection effects through matching, we estimate the wage gap using standard OLS regression, where the depended variable is the logarithm of the absolute yearly wage. Table 4 shows the corresponding results. Regressing the dependent variable only on a dummy ‘*ACADEME*’, which indicates academic scientists by taking on value one (zero otherwise), reveals an average wage gap of 19.75% (column 1).¹⁷ The estimated wage gap increases slightly (but not significantly, $\chi^2(1)=0.58$, $p=0.4469$) to 20.90% once other observable factors are taken into account. As can be expected, wage follows an inverse-U shape through time, and increases in job tenure and hours worked. Wages also increase in taste for business.

While this analysis provides some first insights regarding the size of the wage gap, the unequally distributed characteristics of academic and industrial scientists may confound the estimation. Further, it is implicitly assumed that the covariates have the same effects in academe and industry.¹⁸ To mitigate the first issue, we re-estimate the average wage gap using a matched sample. The second issue is dealt with by additional analyses of the determinants of the wage differential using matched academe-industry pairs as the unit of observation.

Table A4 in the appendix presents the Probit model used for the propensity score matching. For 29 academic scientists no adequate match could be found due to lack of common support. Hence, we exclude them from the following investigation. As this reduction amounts to less than 3 percent of the full sample and the sample mean wage and other controls stay almost the same (no statistically significant changes), the sample reduction does not affect the analysis. No

¹⁷ Marginal effect calculated as $\exp(-0.2196)-1$; analogue for the other marginal effects described in this section.

¹⁸ Columns five and six of Table 6 below show that this assumption does not hold for all covariates.

significant differences remain between the two groups after the matching, as shown by two-sided t -tests on differences in the means of each variable in Table 5. Further, Table A4, column 2, provides the results of a re-estimation of the aforementioned Probit model based on the matched sample of academic and industrial scientists. A χ^2 -test on the joint significance of all explanatory variables indicates that both samples show no significant differences in their covariates anymore ($\chi^2(14)=6.10, p=0.9640$).

After matching the difference between the average annual wage in industry and the average annual wage in academe becomes €11.828 (two-sided t -test on mean differences of wages before and after matching: $t(455)=-2.91, p=0.004$), or a decrease of the wage gap by 20.2 percent (initially €14.816). In relative numbers, the descriptive wage gap decreased by 5.8 percentage points from 28.3 percent to 22.5 percent.¹⁹

To estimate the wage gap after matching, we regress wage on the *ACADEME* dummy. Without the inclusion of other controls, the average wage difference between industry and academe is estimated at 12.92% ($p<0.10$, Table 4, column 3). Controlling for other observables, the wage gap is highly significant at 13.19% (column 4).²⁰

The analysis suggests that descriptive comparisons of average wages and estimates of the wage gap using standard OLS regressions overestimate the actual wage gap. Specifically, the average wage gap narrowed significantly from a 28% difference in means, to about 21% when using standard OLS regression, and further down to 13% if OLS regressions are combined with a balanced sample of academic and industrial scientist after matching. Hence, descriptive mean comparisons overestimate the wage gap by about 15 percentage points.

¹⁹ Note that all results are even more pronounced if one considers the median wages instead of the means.

²⁰ While we find some significant effects of covariates on wage after matching, this does not mean that the matching was unsuccessful. Rather, they should be interpreted as common trends across industry and academe.

Table 4: OLS Regression results

Dependent variable	(1)		(2)		(3)		(4)	
	Before matching				After matching			
			ln(<i>wage</i>)					
	Coef	Se	Coef	Se	Coef	Se	Coef	Se
<i>ACADEME</i>	-0.2196***	(0.0207)	-0.2345***	(0.0202)	-0.1384*	(0.0727)	-0.1415***	(0.0436)
<i>TASTE_SCIENCE</i>			-0.0087	(0.0088)			-0.0210	(0.0229)
<i>RES_SHARE</i>			0.0002	(0.0003)			-0.0010*	(0.0006)
<i>DEV_SHARE</i>			-0.0001	(0.0003)			0.0015*	(0.0009)
<i>TASTE_BUSINESS</i>			0.0140*	(0.0072)			0.0202	(0.0312)
<i>CAREER_YEARS</i>			0.0440***	(0.0042)			0.0695***	(0.0141)
<i>CAREER_YEARS_2</i>			-0.0007***	(0.0001)			-0.0017***	(0.0005)
<i>TIME_CURRENTJOB</i>			0.0054***	(0.0017)			0.0082**	(0.0036)
<i>RELATIONSHIP</i>			0.0522**	(0.0237)			0.1393***	(0.0524)
<i>CHILDREN</i>			0.0250	(0.0189)			-0.0041	(0.0444)
<i>GENDER</i>			-0.0299	(0.0199)			-0.0134	(0.0383)
<i>SCHOLARSHIP</i>			0.0067	(0.0159)			-0.0203	(0.0407)
<i>TIMETOPHD_PROP</i>			-0.0182	(0.0292)			-0.0882	(0.0693)
<i>HOURS_WORKED</i>			0.0084***	(0.0012)			0.0027	(0.0034)
<i>NATURAL</i>			-0.0249	(0.0156)			-0.0737*	(0.0393)
<i>INTERCEPT</i>	11.0298***	(0.0145)	10.1754***	(0.0738)	10.9542***	(0.0711)	10.3544***	(0.1706)
Number of observations	1245		1245		912		912	
<i>R</i> ²	0.076		0.537		0.028		0.488	
Root MSE	0.3727		0.2653		0.4098		0.2997	
<i>F</i> -statistic	112.4988		76.4002		3.6207		13.9787	

Notes: This table presents OLS regressions of wage before and after matching. Heteroscedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$. (3-4): Clustered standard errors take repeated observations caused by sampling with replacement into account. Definitions of the variables are provided in Table 1.

Table 5: Summary statistics after matching

Variable	Academia after matching (n=456)					Selected controls in industry (n=456)					<i>p</i> -value of two-sided <i>t</i> -test on mean differences [#]
	Mean	Std. Dev.	Min	Median	Max	Mean	Std. Dev.	Min	Median	Max	
<i>TASTE_SCIENCE</i>	0.40	0.87	-1.63	1.02	1.27	0.42	0.84	-1.34	1.02	1.47	p=0.830
<i>RES_SHARE</i>	47.78	34.95	0.00	50.00	100.00	50.23	41.80	0.00	52.50	100.0	p=0.619
<i>DEV_SHARE</i>	7.89	16.02	0.00	0.00	100.00	9.37	14.95	0.00	0.00	90.00	p=0.429
<i>TASTE_BUSINESS</i>	-0.11	0.44	-0.55	-0.29	2.05	-0.13	0.47	-0.55	-0.29	7.06	p=0.783
<i>CAREER_YEARS</i>	12.04	8.32	0.00	10.75	29.75	12.24	8.56	0.00	10.63	29.67	p=0.877
<i>AGE</i>	41.79	8.70	27.00	41.00	63.00	42.16	8.56	0.00	1.00	1.00	p=0.755
<i>TIME_CURRENTJOB</i>	8.69	9.16	0.00	5.00	37.00	8.53	7.50	0.00	6.50	30.00	p=0.849
<i>RELATIONSHIP</i>	0.83	0.37	0.00	1.00	1.00	0.82	0.38	0.00	1.00	1.00	p=0.823
<i>CHILDREN</i>	0.57	0.50	0.00	0.00	1.00	0.58	0.49	0.00	1.00	1.00	p=0.836
<i>GENDER</i>	0.22	0.41	0.00	0.00	1.00	0.16	0.36	0.00	0.00	1.00	p=0.199
<i>TIMETOPHD_PROP</i>	0.99	0.30	0.05	0.96	2.40	1.00	0.29	0.43	0.95	2.72	p=0.912
<i>SCHOLARSHIP</i>	0.52	0.50	0.00	1.00	1.00	0.46	0.50	0.00	0.00	1.00	p=0.393
<i>HOURS_WORKED</i>	49.82	8.22	38.00	50.00	70.00	50.51	8.02	38.00	50.00	70.00	p=0.502
<i>NATURAL SCIENCES</i>	0.62	0.49	0.00	1.00	1.00	0.63	0.48	0.00	1.00	1.00	p=0.860
<i>WAGE</i>	52474.41	17382.71	20000.00	49800.00	120000.00	64302.80	33311.52	21339.00	54000.00	200000.00	p=0.004

Notes: #: *p*-values calculated using Lechner's (2001) approximation of standard errors to account for sampling with replacement. 29 observations are excluded from the analysis after matching because the sample is restricted to common support as described in section 3. Definitions of the variables are provided in Table 1.

We can draw some interesting conclusions from our empirical analysis. First, even though differences in research activities and preferences are believed to be the main drivers of the academe-industry wage differential, the wage gap remains significant when these structural differences are controlled for. This could have several reasons. There might be significant differences in the appropriation of the returns to research between industry and academia. Firms are generally more inclined to commercialize findings, while academic institutes are more inclined to publish them. The differing goals of research might lead a researcher in industry to generate more appropriable value than the same researcher in academia. While the empirical analysis controls for research activities, this remains a structural difference between academic and industrial research.

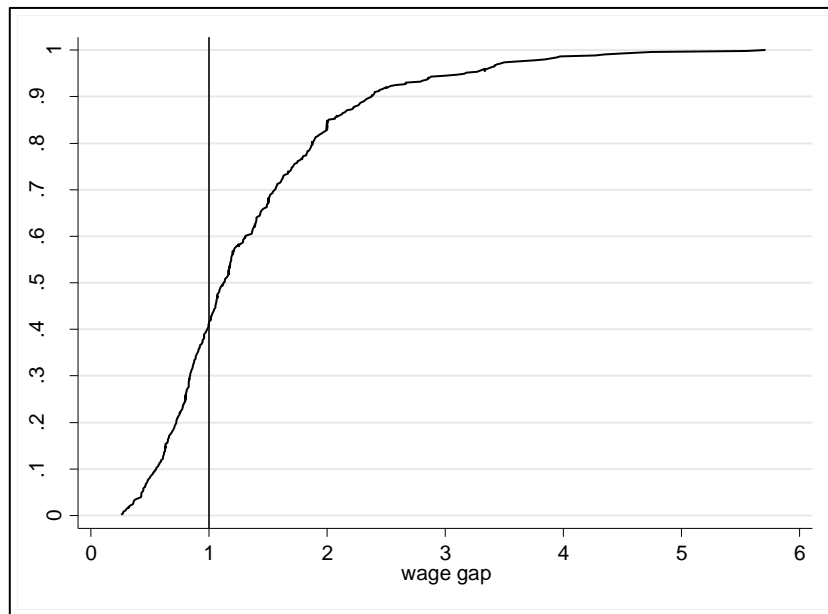
Academic scientists and researchers in industry might also pursue different activities outside of research. Academics mostly teach or participate in the administration of the university when not focusing on research or development, whereas industrial researchers might engage in management activities, which generate significant value for the firm. While we implicitly control for the share of time the researcher spends on non-research activities (through the shares of time spent on research and development), this remains another structural difference.

Last, the expectation that the wage gap closes after matching builds on the belief that industrial scientists are willing to forgo monetary rewards for a higher focus on research activities, while academic scientists' wages are relatively unaffected by the time they spend on research. The relatively small decrease of the wage gap after matching might indicate that other wage determinants that differ systematically across academic and industrial researchers exert a positive influence on the wage differential, such as differences in the time spent on development orientated research activities. In order to disentangle the positive and negative factors underlying

the wage gap, the upcoming section presents regressions of the estimated wage differences on research activities and other covariates.

Figure 2 illustrates the heterogeneity in the wage gap through a plot of the cumulative distribution of the relative difference between the observed wage in academe and the estimated counterfactual wage in industry. While the average descriptive wage gap after matching is 22.5%, 40% of the sampled academics actually earn more than the estimated wage in industry.²¹ On the other hand, almost 20% of the sample might earn at least twice as much if they would work in industry.

Figure 2: Cumulative distribution of the wage gap



²¹ It should be noted here that this evaluation of the wage gap assumes that the academic would work for the average firm, instead of matching to a firm which suits his needs. If this matching were to happen, the wage gap could become even smaller for scientists who spend much time on research, as they are more likely to sort into firms which offer high freedom and lower wages. Scientists who spend more time on development should be more likely to sort into companies which focus more strongly on development, leading to an even higher wage differential.

5.2. Explaining the wage gap

The aim of following analysis is to examine how differences in the time spent on research and development influence the academe-industry wage gap, holding personal characteristics constant. Therefore, we run a standard OLS regression of the estimated wage gap per academic scientist (i.e. $\frac{\text{est. counterfactual wage in industry}}{\text{observed wage in academe}}$) on the scientist's time spent on research, time spent on development, taste for science, taste for business, years since graduation in linear and quadratic terms, time in the current job, demographic controls, controls for ability, average hours worked, and a field control.²² In order to reduce the skewness of the wage gap distribution we estimate a semi-elastic model, taking the natural logarithm of the relative wage differential as the dependent variable. Table 6 presents the results.

²² To avoid multicollinearity we include *CAREER_YEARS* but not *AGE* in the regression. The results are robust to employing *AGE* instead of *CAREER_YEARS*.

Table 6: OLS Regression of the Individual Wage Gap

Dependent variable	(1)		(2)		(3)		(4)		(5)	
	$\ln\left(\frac{\text{Wage in industry}}{\text{Wage in academia}}\right)$		$\ln(wage)$ in Academia		$\ln(wage)$ of selected industry controls					
	Coef	Se	Coef	Se	Coef	Se	Coef	Se	Coef	Se
<i>TASTE_SCIENCE</i>	-0.0162	(0.0273)	-0.0508	(0.0767)	-0.5060	(0.3520)	0.0011	(0.0156)	-0.0396	(0.0360)
<i>RES_SHARE</i>	-0.0032***	(0.0008)	-0.0018	(0.0011)	-0.0046***	(0.0011)	0.0006	(0.0004)	-0.0023***	(0.0008)
<i>DEV_SHARE</i>	0.0030**	(0.0013)	0.0034*	(0.0019)	0.0043**	(0.0020)	0.0001	(0.0006)	0.0028*	(0.0015)
<i>TASTE_BUSINESS</i>	0.0052	(0.0487)	-0.1715*	(0.0988)	-0.0302	(0.0981)	0.0369	(0.0299)	0.0016	(0.0529)
<i>CAREER_YEARS</i>	-0.0315**	(0.0125)	0.0083	(0.0191)	-0.0498***	(0.0159)	0.0309***	(0.0065)	0.0856***	(0.0163)
<i>CAREER_YEARS_2</i>	0.0002	(0.0004)	-0.0009	(0.0006)	0.0009	(0.0006)	-0.0003	(0.0002)	-0.0024***	(0.0006)
<i>TIME_CURRENTJOB</i>	-0.0042	(0.0039)	-0.0010	(0.0057)	-0.0098*	(0.0059)	0.0003	(0.0019)	0.0218***	(0.0065)
<i>RELATIONSHIP</i>	-0.0694	(0.0761)	-0.0722	(0.1196)	-0.0521	(0.0930)	-0.0026	(0.0415)	0.3474***	(0.0913)
<i>CHILDREN</i>	0.0518	(0.0575)	0.0532	(0.0879)	0.0576	(0.0761)	0.0314	(0.0268)	-0.0400	(0.0707)
<i>GENDER</i>	0.0688	(0.0690)	0.2842***	(0.1031)	-0.0713	(0.0885)	-0.0761**	(0.0323)	0.0227	(0.0768)
<i>SCHOLARSHIP</i>	0.1034**	(0.0515)	0.0551	(0.0740)	0.1234*	(0.0718)	-0.0287	(0.0245)	-0.0452	(0.0560)
<i>TIMETOPHD_PROP</i>	-0.0679	(0.0754)	-0.0962	(0.1224)	-0.0747	(0.1018)	-0.0040	(0.0376)	-0.1353	(0.1041)
<i>HOURS_WORKED</i>	-0.0055*	(0.0031)	0.0003	(0.0041)	-0.0101**	(0.0042)	0.0053***	(0.0015)	-0.0002	(0.0047)
<i>NATURAL</i>	0.0260	(0.0513)	0.0890	(0.0742)	-0.0042	(0.0688)	-0.0621***	(0.0235)	-0.0758	(0.0634)
<i>INTERCEPT</i>	0.9240***	(0.2195)	0.2172	(0.3044)	1.8725***	(0.4451)	10.2601***	(0.1159)	10.3016***	(0.2592)
Number of observations	456		174		282		456		456	
R^2	0.226		0.254		0.280		0.473		0.628	
Root MSE	0.5170		0.4645		0.5357		0.2389		0.2977	
F -statistic	10.6699		4.8140		8.5379		29.6733		11.4601	

Notes: This table presents OLS regressions of the wage gap as generated by the matching procedure. Heteroscedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *, $p < 0.10$; **, $p < 0.05$; ***, $p < 0.01$. (1) dependent variable: treatment effect ($\ln(\text{wage in industry}/\text{wage in academe})$), full sample. (2) treatment effect regression for scientists with taste for science lower than or equal to the mean. (3) treatment effect regression for scientists with taste for science larger than the mean. (4) dependent variable: $\ln(\text{wage})$ of academic scientists. (5) dependent variable: $\ln(\text{wage})$ of selected industrial researchers. Clustered standard errors take repeated observations caused by sampling with replacement into account. Definitions of the variables are provided in Table 1.

The evidence supports the hypothesis that the wage gap is smaller for scientists who spend more time on research: it declines by 0.32 percent for each additional percent of time spent on research (column 1). Consistent with previous findings (e.g. Stern, 2004) and the notion of a dominating preference effect (hypothesis 1), differences in research orientation between academic and industrial scientists explain a large part of the positive wage differential. Comparing only academic scientists with those industrial scientists that are able to spend similar amounts of time on research, leaving all other covariates unaccounted, the academe-industry wage gap would indeed be much smaller than descriptive statistics suggest. Academic scientists who are full-time engaged in research would earn only 1.6% more in industry. In line with our conceptualization of wage dynamics in industry and academe, this effect is driven by a negative relation between research and wage in industry, while academic wages are not sensitive to differences in research time (models 4 and 5, Table 5).

Spending more time on development increases the wage gap. With each additional percent of time spent on development oriented research by academic scientists, a comparable industrial scientist's wage increases by 0.3 percent. The magnitude of the effect is not trivial: an average academic scientist who spends 50 percent of his time on development oriented research is subject to an estimated wage differential of 73.5 percent.^{23,24} Wage regressions within both sectors (models 4 and 5, Table 5) further show that this effect stems largely from a positive relation between development oriented research and wages within industry, while the relation is insignificant within academe. In terms of preference versus productivity effects, the negative relation between research and wages can be interpreted as preference effects dominating productivity effects, in line with the findings of Stern (2004) and Sauermann and Roach (2014).

²³ 23 academics in the sample, or 5%, spend at least 50% of their time on development.

²⁴ As noted above, these predictions assume that scientists would work for the average firm, instead of sorting into firms according to their preferences.

That development relates positively to industrial wages could either indicate a higher productivity effect, i.e. there occurs more rent-sharing with scientists who focus on more profitable development, or a lower preference effect, scientists enjoy working on development less and demand higher wages to do so. While our data do not allow us to separate these alternatives, we do show that the relation between wage and science is not exclusively negative: research in its most applied form relates to higher wages in industry.

Apart from differences in research activities, the wage gap is also strongly affected by time since graduation: the wage gap decreases by approximately 3.2 percent per year.²⁵ This result is in line with earlier findings of Stevens (2004) and Agarwal and Ohyama (2013), who both report converging earnings paths of academic and industrial scientists in the US. The effect seems to be driven by diminishing wage increases throughout time in industry, compared to linear wage increments in academe (column 4 and 5, Table 5).

To test the moderating role of taste for science on the relation between research orientation and the wage gap (hypothesis 3), we split the sample in a low-taste-for-science and high-taste-for-science group based on the mean values (columns 2 and 3).²⁶ In line with the conceptual reasoning, the impact of *RES_SHARE* on the wage differential is higher in the high-taste-for-science group (column 3) than in the low-taste-for-science group (column 2), indicating that high taste for science researchers are willing to trade of wage for more research orientation at higher levels than low taste for science researchers (*F*-test on differences between the coefficients of *RES_SHARE* in model 2 and 3: $F(1,426)=3.12$, $p=0.0780$). The results imply that academic

²⁵ Since the turning point of the estimated relationship between career years and the relative wage gap lies far out of the observed range of career years (315 years), the influence is almost linear with only slightly decreasing marginal effects. Including only the linear term or taking a logarithmic specification does not alter the results.

²⁶ Alternatively, one could include a corresponding interaction term in the regression. However, interaction term specifications imply the assumption that all other covariates exert the same effect within both sub-groups, which is in our case not true.

researchers with a relatively high taste for science who spend much time on research face much lower comparable wages in industry. All other covariates equal, the wage gap for scientists with a rather high taste for science becomes zero at 86 percent of time spent on research. Scientists with a high taste for science who want to spend most of their time on research might thus be financially best off in an academic position. The effect of time spent on development is also larger for scientists with a high taste for science compared to those with a low taste for science, but the difference is not statistically significant (0.0034 in the low group versus 0.0043 in the high group, $F(1,426)=0.11$, $p=0.7460$).

One limitation of this analysis is that the measures for research and development activities are rather broad and the actual underlying activities might still be quite heterogeneous. Even though they are more detailed than other measures used in previous studies, even more detailed information is needed to further disentangle which kinds of research activities, such as the possibility to publish research findings, are mainly responsible for wage differences between both sectors. ‘Research’ and ‘Development’ might also carry different meanings in academe and industry. While the survey clearly stated what is meant by each activity, divergent interpretations between the sectors remain a concern.

6. Conclusion

This paper provides new evidence on the academe-industry wage gap. Once common selection effects in the scientific labor market and potential confounding personal characteristics have been accounted for, the estimated wage gap is about 15 percentage points smaller than descriptive statistics would suggest. The wage differential varies with individual scientists’ focus on research versus development activities: academic scientists who spend much time on research

face much lower counterfactual wages, whereas academic scientists who spend much time on development activities face much higher ones. How much scientists pay to be scientists thus hinges on their research orientation.

Further, the trade-off between wage and research time is much stronger for scientists with an above-average taste for science. Combined with the negative relationship between wages and research time in industry and the positive relationship between wages and development time, it could be argued that in terms of ‘research’ activities preference effects dominate productivity effects, whereas in terms of ‘development’ activities productivity effects dominate preference effects.

While self-selection and different personal characteristics of academic and industrial scientists explain a large part of the wage gap, some difference still remains. Probable reasons for the remaining wage gap are a better appropriation and monetization of research findings in industry, differences in non-research related activities among academics and researchers, e.g. teaching versus managing, and the noncompetitive nature of academic wages in Belgium.

Our results have practical implications for academic institutions, policy makers, and individual scientists. For academic institutions it is worthwhile to note that offering scientists the freedom to pursue their own research projects compensates for much lower monetary rewards in academe. Urging scientists to focus on more applied research projects might cause selection of academics into industry, since scientists can expect to receive much higher returns to development oriented research in industrial positions. This selection might be unwanted from the perspective of the university if it causes better scientists to leave academe.²⁷ On the other hand, allowing for more commercialization might also create a more enjoyable academic workplace for

²⁷ In that regard, Balsmeier and Pellens (2014) find that academic scientists that patent are more likely to leave academe, and also publish more than average.

scientists who would have otherwise preferred a career in industry. The net effect of increased commercialization of research findings on the quality of academic scientists remains unknown but might be addressed by future research.

If the obligatory commercialization of academic scientists' research findings turns out to be a source of academic brain drain, it might be favorable to alleviate it. If commercialization is preferred, academic institutes might also put more weight on successful research commercialization when it comes to promotion decisions. That way, engaging in applied research can have positive wage benefits in academe, which would narrow the wage gap.

From a policy point of view the relatively low wage gap faced by research focused scientists and the rather large wage gap of development oriented scientists might ensure an efficient allocation of different types of scientists into both sectors. Politicians and managers of academic institutions should, however, be aware of the fact that the less time academic scientists are able to focus on research the more attractive it becomes for scientists to seek an industrial position.

It is imperative to keep in mind that these results come from a setting in which academic wages are subject to strong regulation. This is likely to affect the estimates of the size of the wage gap but is less likely to affect our assessment of selection and wage-research relationships, as the latter are very much in line with relationships found in US-based data.²⁸ We can only provide some reasonable speculation about how the wage gap in more competitive environments might relate to our calculations. Assuming that such a system leads to higher wage offers (or higher salary due to performance contingent pay) for highly performant academic researchers, the industry-academe wage gap could become smaller for highly able academic scientists who

²⁸ For instance, our findings regarding motive-based selection into industry or academe are similar to those of Roach and Sauermann (2010), Sauermann and Roach (2012) and Agarwal and Ohyama (2013). Sauermann and Roach (2014) uncovers a similar relationship between motives and reservation wages for publishing.

focus on research. They might even experience a wage premium compared to what they could have earned in industry conditional on similar activities. The wage gap might also become wider for lowly skilled academics who focus on research, since they might attract lower wage offers than the academics in our sample. Differences in the relation between commercialization and the wage gap critically depend on whether commercialization performance is part of the wage determination. If it is, the wage gap might widen less strongly or might even become narrower for highly skilled scientists. Further analyses are required to test these notions empirically.

Our findings also have a bearing on individual researchers' decisions whether to pursue a career in academe or industry. As getting into the academic labor market becomes harder, scientists could consider employment in research-like industrial jobs. While those jobs offer lower wages compared to development-focused industrial research jobs, the wage difference compared to academe turns out to be rather small. Hence, they might be considered as welcome outlets for scientists who are forced to leave academe but still want to pursue research. It is also interesting to know that the perceived higher wages in industry are conditional on a focus on development oriented research and other activities that are unrelated to research. Future investigations might be able to answer the question which role the individual wage gap plays for scientists' career decision and how it influences the allocation of research talent into both sectors.

References

- Agarwal, R., Ohyama, A. (2013). Industry or academia, basic or applied? Career choices and earnings trajectories of scientists. *Management Science*, 59, 950-970.
- Aghion, P., Dewatripont, M., Stein, J. (2008). Academic freedom, private-sector focus, and the process of innovation. *Rand Journal of Economics*, 39(3), 617-635.
- Altbach, P., Reisberg, L., Yudkevich, M., Androushcak, G., Pacheco, I. (2012). Paying the professoriate: a global comparison of compensation and contracts. London: Routledge.
- Angrist, J. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, 66(2), 249-288.
- Arora, A., Gambardella, A. (1994). The changing technology and technological change: general and abstract knowledge and the division of innovative labour. *Research Policy*, 23(5), 523-532.
- Auriol, L. (2007). Labour market characteristics and international mobility of doctorate holders: results for seven countries. *OECD Science, Technology and Industry Working Papers* 2007/02. Retrieved from <http://dx.doi.org/10.1787/310254328811>.
- Auriol, L. (2010). Careers of doctorate holders: employment and mobility patterns. *OECD Science, Technology and Industry Working Papers* 2010/04. Retrieved from <http://dx.doi.org/10.1787/5kmh8phxvfv5-en>.
- Auriol, L., Schaaper, M., Felix, B. (2012). Mapping careers and mobility of doctorate holders: draft guidelines, model questionnaire and indicators. *OECD Science, Technology and Industry Working papers* 2007/6. Retrieved from <http://dx.doi.org/10.1787/5k4dnq2h4n5c-en>.

- Balsmeier, B., Pellens, M. (2014). Who makes, who breaks: which scientists stay in academe? *Economics Letters* 122(2), 229-232.
- Barbezat, D. (1987). Salary differentials by sex in the academic labour market. *Journal of Human Resources*, 22, 422-428.
- Bayer, A., Astin, H. (1968). Sex differences in academic rank and salary among science doctorates in teaching. *Journal of Human Resources*, 2, 191-200.
- Blundell, R., Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconometrics. *Journal of Human Resources*, 44, 565-640.
- Card, D., Sullivan, D. (1988). Measuring the effect of subsidized training-programs on movements in and out of employment. *Econometrica*, 56, 497-530.
- Cohen W.M., & Levinthal D.A. (1989). Innovation and learning: the two faces of R&D. *The Economic Journal* 99(397), 569-596.
- Cohen, W., Levinthal, D. (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128-152.
- Creedy, J. (1988). Cohort and cross-sectional earnings profiles: scientists in Britain and Australia. *Journal of Economic Studies*, 15(1), 44-52.
- Dasgupta, P., David, P. (1994). Towards a new economics of science. *Research Policy*, 23, 487-521.
- Diamond, A. (1986). The life-cycle research productivity of mathematicians and scientists. *Journal of Gerontology*, 41, 520-525.
- Eurostat (2012). Careers of doctorate holders. Retrieved 05 03, 2012, from http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Careers_of_doctorate_holders

- Federaal Wetenschapsbeleid. (2006). Careers of Doctorate Holders Survey [Database]. Gent: ECOOM UGent.
- Fini, R., Lacetera, N. (2010). Different yokes for different folks: individual preference, institutional logics, and the commercialization of academic research. In G. Libecap (Ed.), *Spanning boundaries and disciplines: university technology commercialization in the idea age* (Vol. 21, pp. 1-25). Bingley: Emerald Group Publishing.
- Frolich, M. (2007). Propensity score matching without conditional independence assumption - with an application to the gender wage gap in the United Kingdom. *Econometric Journal*, 10, 359-407.
- FWO. (2013). Ph.D. Fellowship. Retrieved 01 27, 2013, from FWO: www.fwo.be/Aspirant.aspx.
- Gerfin, M., Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. *Economic Journal*, 112, 854-893.
- Heckman, J. (1976). The common structure of statistical models for truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4), 475-492.
- Heckman, J.J., Ichimura, H., Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies* 65(2), 261-294
- Imbens, G., Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47, 5-86.
- Konrad, A., Pfeffer, J. (1990). Do you get what you deserve? Factors affecting the relationship between productivity and pay. *Administrative Science Quarterly*, 35, 258-285.
- Lacetera, N. (2009). Different missions and commitment power in RD organizations: theory and evidence on industry-university alliances. *Organization Science*, 20, 565-582.

- Laitner, J., Stafford, F. (1985). The academic labor market: has compensation diverged from other professions? *Econometric Society Meetings*. Washington, DC.
- Lechner, M. (2001), Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, in: M. Lechner and F. Pfeiffer (eds.), *Econometric evaluation of active labor market policies*, Heidelberg, 43-58.
- Lillard, L., Weiss, Y. (1979). Components of variation in panel earnings data: American scientists. *Econometrica*, 47, 437-454.
- McNabb, R., Wass, V. (1997). Male-female salary differentials in British universities. *Oxford Economic Papers*, 49, 328-343.
- Merton, R. (1973). The sociology of science: theoretical and empirical investigations. Chicago: University of Chicago Press.
- Moore, W., Newman, R., Turnball, G. (1998). Do academic wages decrease with seniority? *Journal of Labour Economics*, 16, 352-366.
- Moortgat, P., Van Mellaert, G. (2011). CDH (Careers of Doctorate Holders). Onderzoek, ontwikkeling en innovatie in België studiereeks 12 (Available at http://www.belspo.be/belspo/organisation/Publ/pub_ostc/ind/ind12_nl.pdf ed.). Brussels, Belgium: Belgian Science Policy Office.
- National Science Board. (2012). Science and Engineering Indicators 2012. Arlington VA: National Science Foundation.
- Nopo, H. (2008). Matching as a tool to decompose wage gaps. *Review of Economics and Statistics*, 90, 290-299.
- OECD (2013). OECD/UNESCO Institute for Statistics/Eurostat Careers of Doctorate Holders (CDH) project. Retrieved 08 06, 2013, from

<http://www.oecd.org/innovation/inno/oecdunescoinstituteforstatisticseurostatcareersofdoctorateholderscdhproject.htm>

- Ong, L., Mitchell, J. (2000). Professors and hamburgers: an international comparison of relative academic salaries. *Applied Economics*, 32, 869-876.
- Roach, M., Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. *Research Policy*, 39, 422-434.
- Rosenbaum, P., Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, 19, 165-174.
- Rubin, D. (1977). Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.
- Sauermann, H., Roach, M. (2012). Taste for science, taste for commercialization, and hybrid scientists. *Paper presented at DRUID conference*, June 19-21, Copenhagen Business School, Denmark.
- Sauermann, H., Roach, M. (2014). Not all scientists pay to be scientists: heterogeneous preferences for publishing in industrial employment. *Research Policy*, 43(1), 32-47.
- Sauermann, H., Stephan, P. (2013). Conflicting logics? A multidimensional view of industrial and academic science. *Organization Science* 24(3), 889-909.
- Stephan, P., Levin, S. (1992). *Striking the mother lode in science: the importance of age, place, and time*. Oxford: Oxford University Press.
- Stephan. (1996). The economics of science. *Journal of Economic Literature*, 34, 1199-1235.
- Stern, S. (2004). Do scientists pay to be scientists? *Management Science*, 50, 835-853.

Stevens, P. (2004). Academic salaries in the UK and US. *National Institute Economic Review*, 190, 104-113.

Uebersax, J. (2000). Estimating a latent trait model by factor analysis of tetrachoric correlations. Retrieved 03 16, 2011, from <http://john-uebersax.com/stat/irt.htm>.

Unesco (2012). Tracking the careers of doctorate holders. Retrieved 07 05, 2012, from <http://www.uis.unesco.org/ScienceTechnology/Pages/doctorate-degree-holders.aspx>

Walker, J., Vignoles, A., Collins, M. (2010). Higher education academic salaries in the UK. *Oxford Economic Papers*, 62, 13-35.

Weiss, Y., Lillard, L. (1978). Experience, vintage, and time effects in the growth of earnings: American scientists. *Journal of Political Economy*, 86, 427-474.

Appendix 1: Data

Starting from the initial dataset of the Careers of Doctorate Holders (CDH) survey we discarded a number of observations for various reasons. Of the 7,160 responses, 5,601 were employed at the time of the survey. 1,973 of those employed were employed in sectors other than academe or the private sector (such as the public sector or private nonprofits). Of the 3,628 remaining observations, 1,713 were active in fields other than natural sciences or engineering.

The remaining 1,915 observations were further scrutinized for inconsistencies.²⁹ 53 academic scientists were removed from the sample because they reported distinctly non-academic job descriptions, including ‘webmaster’, ‘secretary’, ‘manager’, or ‘coordinator’. Respondents who reported being employed in multiple positions were removed from the sample. These are problematic because job characteristics are only reported for the main position, and because it is not obvious to assign scientists with academic and industrial positions to one of the two groups. 38 academic scientists and 28 industrial researchers were thus dropped from the sample. Note that the academic scientists were mainly removed because of industry involvement, while the industrial researchers were employed in multiple jobs in industry. Also note that respondents (especially academic scientists) might still hold multiple affiliations on paper if they did not report them. As such, this restriction serves to remove respondents who in practice hold multiple jobs.

Those who did not report wage or reported zero wages (224 observations), wages below the Belgian annual minimum wage (20 observations) or above the 99th percentile (€200.000, 16 observations) were not included in the analysis. Wages below the minimum wage are either

²⁹ The groups described here can overlap, e.g. they were all retained in the determination of which scientists to remove from the sample and then removed all the same time.

misreported (caused, for instance, by reporting monthly income instead of annual income), or the result of part-time employment, which is removed in the next step. The last group is so rare and unique in the Belgian academic labor market that it would be meaningless to search for comparable scientists. They also have unwarrantedly strong influence on the average wages. These limitations are however not likely to exclude specific groups of scientists (for instance, lower-earning females) from the analysis.

In order to further ensure a reasonable comparability of the scientists included in the analysis, the sample was further restricted to scientists and researchers in the first 30 years after graduation (removing 232 scientists) who were older than 18 and younger than 65 at the moment the survey was carried out (removing 9 scientists). We also removed scientists and researchers who reported working less than full-time (38 hours, 86 observations) or more than 75 hours a week (the 99th percentile, 21 observations). After further removing missing values on the variables of interest, this resulted in a total sample of 1245 scientists, 486 of which were employed in academe and 759 in industry.

Lastly, scientists and researchers who reported spending in total more than 100 percent of their time on research and development were re-scaled to reflect a maximum time allocation of 100 percent.

Appendix 2: Exploratory Factor Analysis used in Tastes Calculation

Since the items are all binary, we employ the Tetrachoric correlation matrix for this analysis (Uebersax, 2000).³⁰ The analysis returns a two factor solution, explaining 87 percent of total variation. Several rotations were applied from which we select the Varimax solution for interpretability. The first factor correlates with motivation through salary, extralegal benefits, career prospects, job security, and weakly with work circumstances. We name this factor “Taste for Business”. The second factor correlates with motivation through intellectual challenge, independence, contribution to society, and, more strongly, with work circumstances. We name this factor “Taste for Science”. Both were normalized for interpretability.

Table A1: Tetrachoric correlation matrix

	1	2	3	4	5	6	7	8
<i>INTELLECTUAL CHALLENGE</i>	1.00							
<i>SALARY</i>	0.24	1.00						
<i>EXTRALEGAL BENEFITS</i>	0.19	0.73	1.00					
<i>CAREER</i>	0.42	0.47	0.36	1.00				
<i>JOB SECURITY</i>	0.25	0.58	0.70	0.45	1.00			
<i>WORK CIRCUMSTANCES</i>	0.42	0.41	0.42	0.23	0.33	1.00		
<i>INDEPENDENCE</i>	0.79	0.13	-0.04	0.18	0.28	0.51	1.00	
<i>CONTRIBUTION TO SOCIETY</i>	0.59	0.18	0.34	0.13	0.36	0.21	0.45	1.00

Table A2: Results of EFA

Item	Factor loadings		Uniqueness
	Taste for business	Taste for science	
<i>INTELLECTUAL CHALLENGE</i>	0.17	0.89	0.18
<i>SALARY</i>	0.78	0.12	0.38
<i>EXTRALEGAL BENEFITS</i>	0.92	0.02	0.16
<i>CAREER</i>	0.51	0.24	0.68
<i>JOB SECURITY</i>	0.77	0.22	0.35
<i>WORK CIRCUMSTANCES</i>	0.40	0.47	0.62
<i>INDEPENDENCE</i>	0.00	0.92	0.15
<i>CONTRIBUTION TO SOCIETY</i>	0.25	0.54	0.65
Eigenvalue	2.55	2.28	
% of variance	46%	41%	

Note: Varimax rotation. Factor loadings over .40 appear in bold.

³⁰ Similar methods have been employed by Sauermann & Roach (2012).

Table A3: Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 <i>TASTE_SCIENCE</i>	1.00															
2 <i>RES_SHARE</i>	0.24*	1.00														
3 <i>DEV_SHARE</i>	-0.15*	-0.23*	1.00													
4 <i>TASTE_BUSINESS</i>	0.03	0.01	0.02	1.00												
5 <i>CAREER_YEARS</i>	0.01	-0.12*	-0.12*	0.07*	1.00											
6 <i>AGE</i>	0.00	-0.13*	-0.11*	0.05	0.95*	1.00										
7 <i>TIME_CURRENTJOB</i>	-0.03	-0.07*	-0.07*	0.06*	0.73*	0.73*	1.00									
8 <i>RELATIONSHIP</i>	0.00	0.04	-0.04	0.03	0.08*	0.07*	0.05	1.00								
9 <i>CHILDREN</i>	0.02	-0.01	0.03	0.02	-0.03	-0.02	-0.06*	0.38*	1.00							
10 <i>GENDER</i>	-0.03	0.07*	-0.13*	-0.05	0.13*	-0.14*	-0.08*	-0.11*	-0.15*	1.00						
11 <i>TIMETOPHD_PROP</i>	-0.01	-0.04	-0.01	-0.03	-0.02	0.09*	0.06*	0.01	0.00	0.03	1.00					
12 <i>SCHOLARSHIP</i>	0.03	0.07*	0.01	0.00	0.04	-0.06*	-0.05	0.03	0.01	0.02	-0.16*	1.00				
13 <i>HOURS_WORKED</i>	0.04	0.01	-0.06*	0.02	0.2*	0.21*	0.16	0.10*	0.12*	-0.20*	-0.01	-0.02	1.00			
14 <i>NATURAL SCIENCES</i>	0.00	0.07*	-0.10*	0.04	0.11*	0.06*	0.03	-0.04	-0.03	0.07*	-0.02	0.07*	-0.09*	1.00		
15 <i>WAGE</i>	-0.11	-0.14*	0.07*	0.11*	0.59*	0.56*	0.46*	0.13*	0.07*	-0.16*	-0.04	0.02	0.27*	-0.01	1.00	
16 <i>HIGHER</i>	0.33*	0.28*	-0.50*	-0.09*	0.04	0.05	0.08*	-0.02	-0.07*	0.07*	0.03	-0.04	0.11*	0.10*	-0.27*	1.00

Notes: *: significant at $p < 0.05$. Variable definitions are provided in Table 1.

Table A4: Propensity score regressions

Dependent variable: Treatment group	(1)		(2)	
	Before matching		After matching	
<i>TASTE_SCIENCE</i>	0.4295***	(0.0478)	0.0031	(0.0753)
<i>RES_SHARE</i>	0.0071***	(0.0013)	-0.0012	(0.0022)
<i>DEV_SHARE</i>	-0.0314***	(0.0021)	-0.0033	(0.0035)
<i>TASTE_BUSINESS</i>	-0.2121***	(0.0763)	0.0562	(0.1202)
<i>CAREER_YEARS</i>	-0.0754***	(0.0204)	-0.0009	(0.0310)
<i>AGE</i>	0.0495**	(0.0197)	-0.0078	(0.0277)
<i>TIME_CURRENTJOB</i>	0.0294***	(0.0085)	0.0088	(0.0123)
<i>RELATIONSHIP</i>	-0.0579	(0.1325)	0.1077	(0.2062)
<i>CHILDREN</i>	-0.2045**	(0.0982)	-0.0204	(0.1594)
<i>GENDER</i>	0.0657	(0.1159)	0.2338	(0.1720)
<i>SCHOLARSHIP</i>	-0.1285	(0.0945)	0.1338	(0.1508)
<i>TIMETOPHD_PROP</i>	0.0110	(0.1676)	0.0426	(0.2512)
<i>HOURS_WORKED</i>	0.0227***	(0.0060)	-0.0040	(0.0107)
<i>NATURAL</i>	0.2867***	(0.0930)	-0.0406	(0.1491)
<i>INTERCEPT</i>	-2.4178***	(0.6555)	0.3478	(1.0027)
N	1245		914	
Pseudo-R ²	0.38		0.01	
Test of joint significance	$\chi^2(14)=630.52***$		$\chi^2(14)=6.10$	

Notes: This table reports Probit model estimations of being in the treatment group. Standard errors in parentheses. After matching: clustered standard errors to account for repeated observations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variable definitions are provided in Table 1.

Appendix 4: Hourly wage results

Table A5: OLS Regression results, hourly wages

Dependent variable	(1)		(2)		(3)		(4)	
	Before matching				After matching			
	Coef	Se	Coef	Se	Coef	Se	Coef	Se
<i>ACADEME</i>	-0.2569***	(0.0198)	-0.2324***	(0.0201)	-0.1242	(0.0830)	-0.1417***	(0.0430)
<i>TASTE_SCIENCE</i>			-0.0092	(0.0088)			-0.0216	(0.0228)
<i>RES_SHARE</i>			0.0002	(0.0003)			-0.0010*	(0.0006)
<i>DEV_SHARE</i>			-0.0001	(0.0003)			0.0015*	(0.0009)
<i>TASTE_BUSINESS</i>			0.0140*	(0.0071)			0.0213	(0.0305)
<i>CAREER_YEARS</i>			0.0438***	(0.0042)			0.0691***	(0.0140)
<i>CAREER_YEARS_2</i>			-0.0007***	(0.0001)			-0.0017***	(0.0005)
<i>TIME_CURRENTJOB</i>			0.0054***	(0.0016)			0.0081**	(0.0035)
<i>RELATIONSHIP</i>			0.0498**	(0.0236)			0.1354***	(0.0522)
<i>CHILDREN</i>			0.0249	(0.0188)			-0.0006	(0.0440)
<i>GENDER</i>			-0.0292	(0.0198)			-0.0106	(0.0377)
<i>SCHOLARSHIP</i>			0.0062	(0.0158)			-0.0174	(0.0402)
<i>TIMETOPHD_PROP</i>			-0.0172	(0.0289)			-0.0812	(0.0683)
<i>HOURS_WORKED</i>			-0.0114***	(0.0012)			-0.0169***	(0.0034)
<i>NATURAL</i>			-0.0228	(0.0155)			-0.0690*	(0.0392)
<i>INTERCEPT</i>	3.2953***	(0.0136)	3.3929***	(0.0729)	3.1737***	(0.0817)	3.5600***	(0.1684)
Number of observations	1245		1245		912		912	
R^2	0.112		0.510		0.022		0.496	
Root MSE	0.3534		0.2640		0.4106		0.2971	
<i>F</i> -statistic	168.3626		74.0830		2.2386		14.6211	

Notes: This table presents OLS regressions of log(wage) before and after matching. Heteroscedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$. (3-4): Clustered standard errors take repeated observations caused by sampling with replacement into account. Definitions of the variables are provided in Table 1.

Table A6: OLS Regression of individual wage gap, hourly wages

Dependent variable	(1)		(2)		(3)		(4)		(5)	
			$\ln\left(\frac{Wage\ in\ industry}{Wage\ in\ academia}\right)$				$\ln(wage)$ in academia		$\ln(wage)$ of selected industry controls	
	Coef	Se	Coef	Se	Coef	Se	Coef	Se	Coef	Se
<i>TASTE_SCIENCE</i>	-0.0455*	(0.0269)	-0.0580	(0.0743)	-0.5824*	(0.3501)	-0.0003	(0.0156)	-0.0409	(0.0356)
<i>RES_SHARE</i>	-0.0040***	(0.0008)	-0.0022*	(0.0012)	-0.0059***	(0.0011)	0.0006	(0.0004)	-0.0023***	(0.0008)
<i>DEV_SHARE</i>	0.0040***	(0.0013)	0.0035*	(0.0018)	0.0060***	(0.0019)	0.0001	(0.0006)	0.0027*	(0.0015)
<i>TASTE_BUSINESS</i>	0.0334	(0.0499)	-0.1576*	(0.0894)	0.0024	(0.0990)	0.0381	(0.0297)	0.0011	(0.0521)
<i>CAREER_YEARS</i>	-0.0303**	(0.0125)	0.0069	(0.0185)	-0.0491***	(0.0159)	0.0304***	(0.0066)	0.0856***	(0.0161)
<i>CAREER_YEARS_2</i>	0.0003	(0.0004)	-0.0009*	(0.0006)	0.0010*	(0.0006)	-0.0002	(0.0002)	-0.0023***	(0.0006)
<i>TIME_CURRENTJOB</i>	-0.0067*	(0.0039)	-0.0003	(0.0052)	-0.0149**	(0.0061)	0.0002	(0.0019)	0.0214***	(0.0064)
<i>RELATIONSHIP</i>	-0.0452	(0.0769)	-0.0681	(0.1176)	-0.0087	(0.0924)	-0.0026	(0.0417)	0.3402***	(0.0910)
<i>CHILDREN</i>	0.0554	(0.0580)	0.0536	(0.0818)	0.0775	(0.0775)	0.0322	(0.0267)	-0.0378	(0.0701)
<i>GENDER</i>	0.0604	(0.0699)	0.2620**	(0.1007)	-0.0715	(0.0903)	-0.0740**	(0.0322)	0.0184	(0.0757)
<i>SCHOLARSHIP</i>	0.1396***	(0.0509)	0.0619	(0.0705)	0.1783**	(0.0706)	-0.0289	(0.0245)	-0.0431	(0.0555)
<i>TIMETOPHD_PROP</i>	-0.1008	(0.0764)	-0.1618	(0.1137)	-0.0986	(0.1053)	-0.0021	(0.0373)	-0.1198	(0.1025)
<i>HOURS_WORKED</i>	0.0129***	(0.0031)	0.0198***	(0.0040)	0.0069*	(0.0041)	-0.0142***	(0.0015)	-0.0195***	(0.0047)
<i>NATURAL</i>	-0.0025	(0.0514)	0.0761	(0.0708)	-0.0457	(0.0691)	-0.0605**	(0.0234)	-0.0697	(0.0630)
<i>INTERCEPT</i>	0.0505	(0.2205)	-0.6337**	(0.2934)	1.0918**	(0.4536)	3.4662***	(0.1157)	3.4830***	(0.2571)
Number of observations	456		174		282		456		456	
R^2	0.238		0.304		0.288		0.434		0.647	
Root MSE	0.5158		0.4446		0.5380		0.2388		0.2947	
F-statistic	10.6697		5.8267		8.0369		26.9804		10.0561	

Notes: This table presents OLS regression results. Heteroskedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$. wage = wage per hour. (1) dependent variable: treatment effect ($\ln(wage)$ in industry/ $\ln(wage)$ in academia), full sample. (2) treatment effect regression for scientists with taste for science lower than or equal to mean. (3) treatment effect regression for scientists with taste for science higher than mean. (4) dependent variable: $\ln(wage)$ of academic scientists. (5) dependent variable: $\ln(wage)$ of selected industrial researchers. Clustered standard errors to take repeated observations caused by sampling with replacement into account. Definitions of the variables are provided in Table 1.